인공지능과 스마트도시

서울시 사업체 증감 예측

B611122 컴퓨터공학전공 우인혜
B711018 컴퓨터공학전공 김동현

이제승 교수님

INDEX

1. 주제 및 데이터 선정

2. 데이터 전처리 과정

3. 예측 모델 추출

4. 예측 데이터 시각화

Contents 01

주제 및 데이터 선정

주제 및 데이터 선정

주제

서울시 사업체 증감 예측

주제 및 데이터 선정

서울시 사업체 조사 DB

연도별 서울 소재 종사자 1인 이상 전 사업체 조사
–〉행정동별 전체 사업체 단위로 재가공
(출처)서울시 빅데이터 캠퍼스

한국 행정 구역 분류

연도별 한국 행정 구역 분류 데이터('94~'2013)
연도별 행정동 변화 반영 위해 사용
(출처)서울시 빅데이터 캠퍼스

제 8차, 9차 한국 표준 산업 분류 코드

크게 18개 업종으로 나누어 조사
결측치 제거 시 사용
(출처)서울시 빅데이터 캠퍼스

Contents 02

데이터 전처리 과정

서울시 사업체 조사 DB

업종 결측치 제거

코드	산업분류명
A	농업 및 임업
01	농업
011	작물 재배업
0111	곡물 및 기타 식량작물 재배업
01110	곡물 및 기타 식량작물 재배업
0112	채소, 화훼작물 및 종묘 재배업

⋮

R	기타 공공, 수리 및 개인서비스업
90	하수처리, 폐기물처리 및 청소관련 서비스업
901	하수, 분뇨 및 축산폐기물 처리업
9011	하수 처리업
90110	하수 처리업
9012	분뇨 및 축산폐기물 처리업
90121	분뇨 처리업
90122	축산폐기물 처리업
902	폐기물 수집운반 및 처리업

코드	산업분류명
A	농업 및 임업
B	어 업
C	광 업
D	제조업
E	전기, 가스 및 수도사업
F	건설업
G	도매 및 소매업
H	숙박 및 음식점업
I	운수업
J	통신업
K	금융 및 보험업
L	부동산업 및 임대업
M	사업서비스업
N	공공행정, 국방 및 사회보장행정
O	교육 서비스업
P	보건 및 사회복지사업
Q	오락, 문화 및 운동관련산업
R	기타 공공, 수리 및 개인서비스업

서울시 사업체 조사 DB

업종 결측치 제거

코드	산업분류명
A	농업 및 임업
01	농업
011	작물 재배업
0111	곡물 및 기타 식량작물 재배업
01110	곡물 및 기타 식량작물 재배업
0112	채소, 화훼작물 및 종묘 재배업

⋮

R	기타 공공, 수리 및 개인서비스업
90	하수처리, 폐기물처리 및 청소관련 서비스업
901	하수, 분뇨 및 축산폐기물 처리업
9011	하수 처리업
90110	하수 처리업
9012	분뇨 및 축산폐기물 처리업
90121	분뇨 처리업
90122	축산폐기물 처리업
902	폐기물 수집운반 및 처리업

코드	산업분류명
A	농업 및 임업
B	어 업
C	광 업
D	제조업
E	전기, 가스 및 수도사업
F	건설업
G	도매 및 소매업
H	숙박 및 음식점업
I	운수업
J	통신업
K	금융 및 보험업
L	부동산업 및 임대업
M	사업서비스업
N	공공행정, 국방 및 사회보장행정
O	교육 서비스업
P	보건 및 사회복지사업
Q	오락, 문화 및 운동관련산업
R	기타 공공, 수리 및 개인서비스업

서울시 사업체 조사 DB

행정 구역 분류 통일

2009.7.1	행정구역명	2009.4.1	행정구역명	비 고
11	서울특별시	11	서울특별시	
11060	동대문구	11060	동대문구	
1106081	용신동	1106051	신설동	2009.5.4기준
		1106077	용두동	행정동 통합
1106082	제기동	1106054	제기1동	2009.5.4기준
		1106055	제기2동	행정동 통합
1106083	전농1동	1106056	전농1동	2009.5.4기준
		1106078	전농2동	행정동 통합
1106084	전농2동	1106058	전농3동	2009.5.4기준 명칭변경
1106085	답십리1동	1106060	답십리1동	2009.5.4기준
		1106079	답십리3동	행정동 통합
1106086	답십리2동	1106061	답십리2동	2009.5.4기준
		1106063	답십리4동	행정동 통합
1106087	장안1동	1106065	장안1동	2009.5.4기준
		1106067	장안3동	행정동 통합
1106088	장안2동	1106066	장안2동	2009.5.4기준
		1106068	장안4동	행정동 통합
1106089	이문1동	1106074	이문1동	2009.5.4기준
		1106075	이문2동	행정동 통합
1106090	이문2동	1106076	이문3동	2009.5.4기준 명칭변경

연도별 한국 행정 구역 분류 데이터
-기간별 변경 내용

서울시 사업체 조사 DB

최종 가공

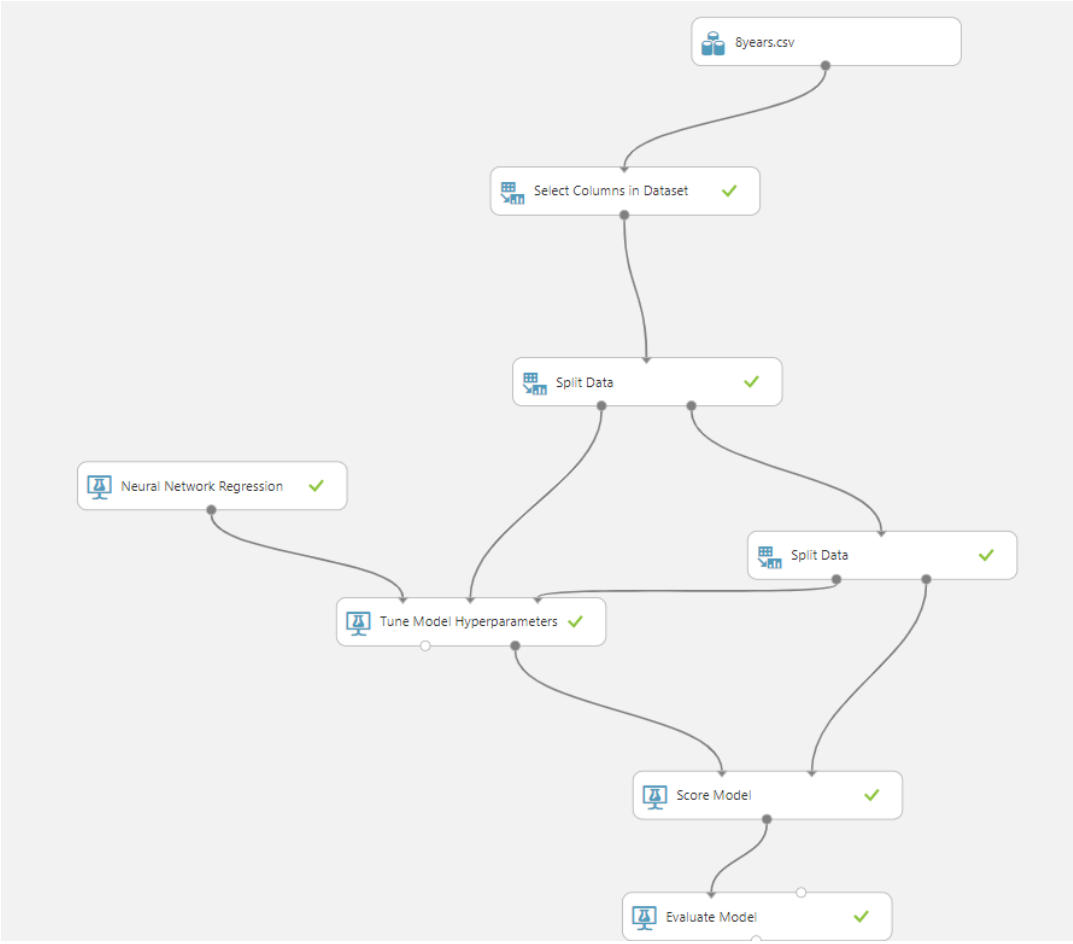
zid	y_1994	y_1995	y_2010	y_2011	y_2012	y_2013
1101053	11.627	12.755	12.831	12.363	12.249	12.79557
1101054	7.51	7.306	7.996	7.652	6.207	6.925033
1101055	6.091	4.708	6.507	7.029	7.613	7.530367
1101056	4.187	4.486	4.49	3.984	4.047	4.14254
1101057	3.792	3.78	2.386	2.329	2.442	2.421937
1101058	7.053	6.492	6.383	8.121	11.262	10.62089
1101060	9.508	11.409	17.46	14.704	10.654	12.24932
1101061	5.168	5.446	5.495	5.337	5.587	5.677058
1101063	3.141	3.278	3.524	3.317	3.186	3.276277
1101064	14.225	15.204	12.681	12.574	12.291	12.76714
1101067	2.955	3.014	2.724	2.579	2.86	2.827336
1101068	2.551	2.542	2.563	2.432	2.726	2.651983
1101069	2.621	2.71	2.609	2.688	2.977	2.933853

Contents 03

예측 모델 추출

예측 모델 추출

Training experiment

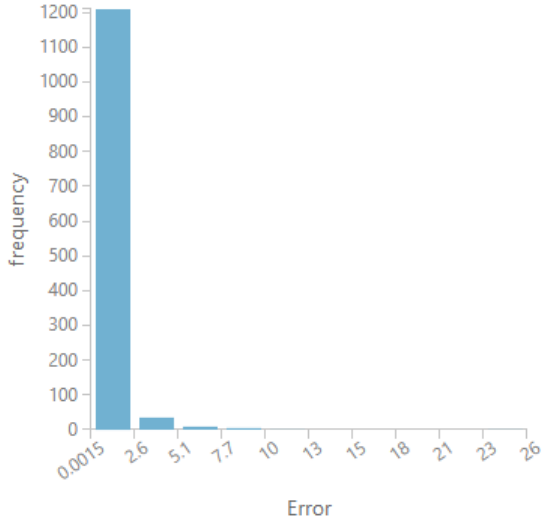


2years › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.645702
Root Mean Squared Error	1.407645
Relative Absolute Error	0.314357
Relative Squared Error	0.196962
Coefficient of Determination	0.803038

Error Histogram



예측 모델 추출

Training experiment

2years › Score Model › Scored dataset

rows 1260 columns 4

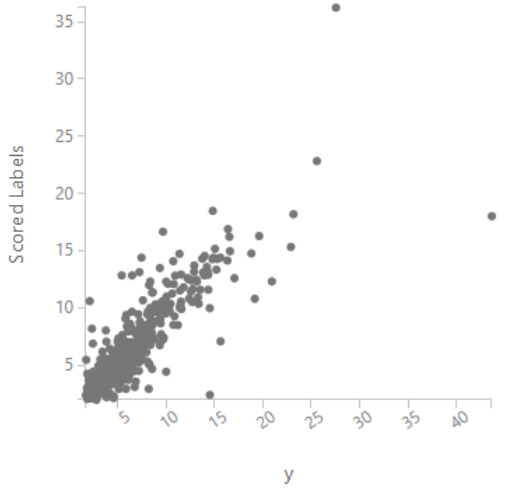
view as

1	2	y	Scored Labels
6.182	6.66	9.348	6.753231
6.04	5.972	6.399	6.119555
6.474	6.742	7.198	6.863584
8.194	6.235	5.854	6.625199
4.357	4.757	4.325	4.843652
2.401	3.718	2.436	3.731263
3.365	3.311	3.298	3.497984
3.148	3.91	3.336	3.977281
2.467	2.346	2.157	2.60337
3.323	3.251	3.278	3.443214
1.722	1.771	1.879	2.058124
3.965	4.681	4.359	4.731442
4.863	4.402	4.788	4.596046
3.155	3.319	3.466	3.481341
16.809	22.892	23.161	18.191214
8.057	4.656	5.983	5.200047
2.028	4.594	4.351	4.430588

Visualizations

y
ScatterPlot

compare to Scored Labels

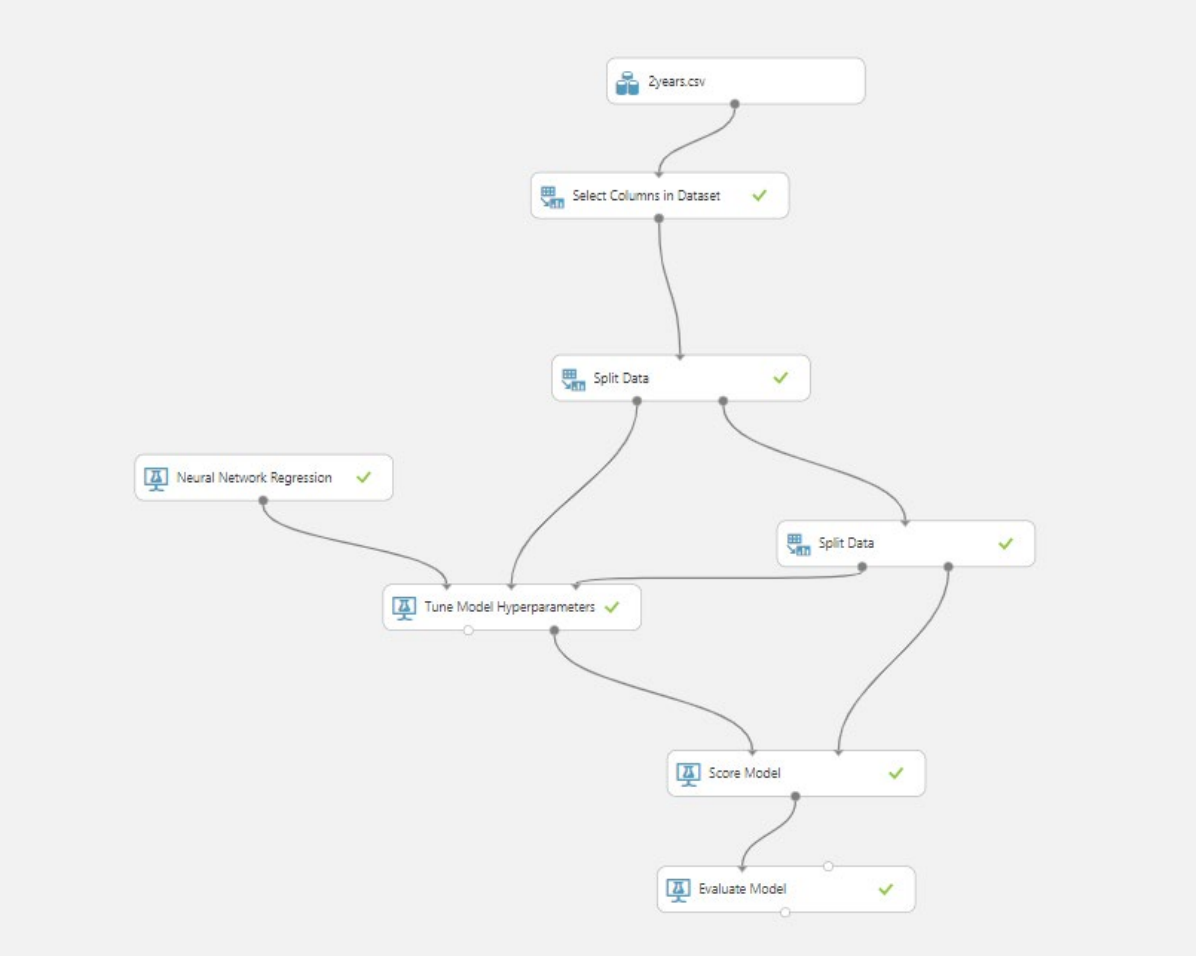


y log scale

Scored Labels log scale

예측 모델 추출

Training experiment

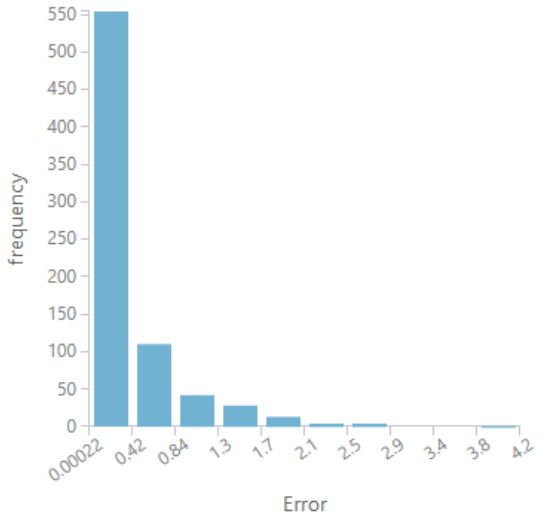


8years › Evaluate Model › Evaluation results

Metrics

Mean Absolute Error	0.374604
Root Mean Squared Error	0.608943
Relative Absolute Error	0.18891
Relative Squared Error	0.048853
Coefficient of Determination	0.951147

Error Histogram



예측 모델 추출

Training experiment

8years › Score Model › Scored dataset

rows 756 columns 10

view as

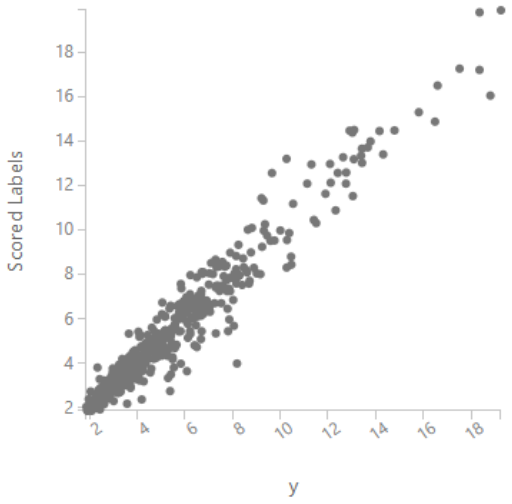
1	2	3	4	5	6	7	8	y	Scored Labels
3.331	3.393	3.246	4.213	5.068	5.61	7.815	5.475	6.336	6.028389
2.237	2.841	2.391	2.065	1.856	2.09	2.181	2.073	2.38	2.116692
2.159	2.557	2.471	2.384	2.379	2.276	2.472	2.413	2.547	2.436646
8.209	6.969	6.999	6.581	6.218	3.991	4.814	6.581	6.56	6.27134
4.156	1.848	2.134	2.124	2.024	2.362	2.26	2.059	2.054	2.116949
3.122	2.531	2.92	2.478	2.749	3.138	2.868	3.054	3.457	3.034168
3.677	3.643	4.109	3.987	3.34	3.315	3.142	2.982	3.064	3.13389
2.218	2.279	3.159	2.94	3.416	3.609	3.559	4.488	4.801	4.25119
3.839	3.92	4.303	4.334	4.093	4.13	4.742	4.768	6.409	4.809353
16.218	16.544	15.055	13.643	13.49	12.892	12.707	12.914	13.389	13.347222
3.985	3.852	3.421	3.748	3.775	3.617	3.616	3.366	3.095	3.488381
2.884	2.848	2.789	2.879	2.918	2.825	2.7	2.605	2.603	2.671775
3.7	4.051	4.093	3.553	4.158	3.635	3.551	3.543	3.586	3.632776
4.044	4.277	5.774	5.216	5.162	4.885	5.359	4.848	4.845	5.126793
7.665	6.849	7.004	6.246	7.756	7.221	7.009	6.277	7.53	6.703289
2.543	2.598	3.874	4.13	4.333	3.927	3.628	3.684	2.982	3.767475
2.564	2.633	2.904	2.977	2.899	2.779	2.717	3.221	3.215	3.110349

Visualizations

y

ScatterPlot

compare to Scored Labels

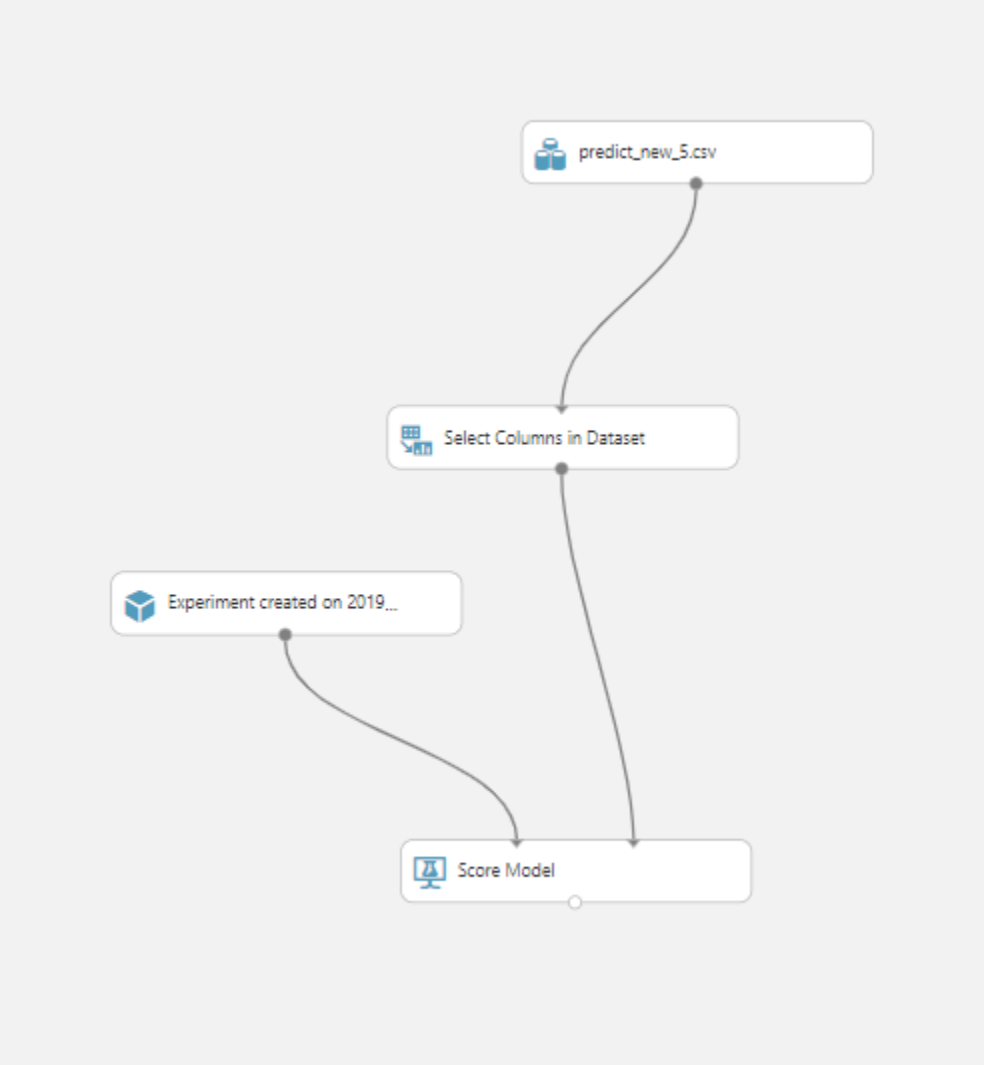


y log scale

Scored Labels log scale

예측 모델 추출

Predictive experiment



예측 모델 추출

Predictive experiment

8years [Predictive Exp.] › Score Model › Scored dataset

rows 420 columns 9

view as

1	2	3	4	5	6	7	8	Scored Labels
11.627	12.755	14.588	13.927	13.025	12.831	12.363	12.249	12.795566
7.51	7.306	8.436	8.736	8.203	7.996	7.652	6.207	6.925033
6.091	4.708	4.297	6.735	6.532	6.507	7.029	7.613	7.530367
4.187	4.486	3.98	4.788	4.592	4.49	3.984	4.047	4.14254
3.792	3.78	2.341	2.488	2.447	2.386	2.329	2.442	2.421937
7.053	6.492	7.289	6.732	7.109	6.383	8.121	11.262	10.620891
9.508	11.409	12.496	12.34	13.688	17.46	14.704	10.654	12.24932
5.168	5.446	5.429	5.884	5.861	5.495	5.337	5.587	5.677058
3.141	3.278	3.274	3.004	3.542	3.524	3.317	3.186	3.276277
14.225	15.204	13.73	15.444	15.607	12.681	12.574	12.291	12.767143
2.955	3.014	3.053	2.806	2.778	2.724	2.579	2.86	2.827336
2.551	2.542	2.267	2.521	2.458	2.563	2.432	2.726	2.651983
2.621	2.71	3.198	2.889	2.615	2.609	2.688	2.977	2.933853
3.284	3.142	3.585	3.066	3.164	3.457	3.314	3.131	3.244653
3.908	4.271	3.973	3.237	3.795	3.816	3.456	3.37	3.486771
6.262	6.281	6.249	6.789	7.604	7.61	6.696	7.244	7.334798
5.49	5.164	5.327	5.576	5.585	5.426	5.408	5.432	5.565952

Statistics

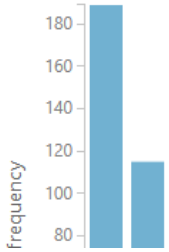
Mean	5.1198
Median	4.1104
Min	1.9437
Max	20.9527
Standard Deviation	2.9533
Unique Values	419
Missing Values	0
Feature Type	Numeric Score

Visualizations

Scored Labels

Histogram

compare to None



Contents 04

예측 데이터 시각화

예측 데이터 시각화

종사자 수 변화율 계산

$$\text{종사자 수 변화율 계산} = \frac{(\text{T+1년도 예측 값}) - (\text{T년도 결과 값})}{(\text{T년도 결과 값})} \times 100$$

https://hongdo.maps.arcgis.com/apps/webappviewer/index.html?id=2e00f48ba85b4b879961addc067489a5

Thank You